Research Paper | Computer Science | E-ISSN No : 2454-9916 | Volume : 3 | Issue : 5 | May 2017

IMPETUS TO DIAGNOSIS IN THE FIELD OF ONCOLOGY WITH THE AID OF DATA MINING APPROACH

Er. Siddharth Arora [1] | Prof. (Dr.) Shiv Kumar Verma [1]

[1] Department of Computer Science & Engineering, Glocal University, State Highway 57, Mirzapur Pole, Saharanpur-247121, Uttar Pradesh, India.

ABSTRACT

The extent of data in the area of real life is escalating with the passage of time. So, to excerpt knowledge from such plenty of data is really very much imperative. So to deal with such a huge data and excerpt knowledge is indeed a very convoluted task. In the area of computer science data mining have a number of techniques to deal with such a plenty of data and provide the fruitful excerpt to the user with only a few effortless steps. Such techniques are pertinent to all the field of science. Various research review had been published regarding the applicability of data mining in assorted field of Sciences as like education, banking, insurance, life science, marketing, telecommunications, medicines etc. For the diagnosis of a diseases a number a variety of distinct test had been suggested from the patient. But by the successful data mining approaches such a number of tests can be curtail. Here in this probe we tried to lend and evaluate how various techniques of data mining can be used for prophecy and diagnosis of dominant cancer affliction.

KEYWORDS: Data mining, Cancer Diagnosis, Cancer Prognosis, Medical Analysis.

I. INTRODUCTION

Cureently, in most of the fields of sciences like education, agriculture, genetics, education, medicine and earth science the amount of data is increasing dramatically. So for the purpose of analysing such a huge plenty of data to excerpt the ideal as well as meaningful information or knowledge discovery is very tedious as well as time consuming deed. So to deal with such kind of situation Data mining techniques are very fruitful for these matters.

Normally, in the field of medical science, there are two phases for taking the decisions. These two decisions phases are as:

1. **Differential Diagnosis:** In the phase of differential diagnosis, whole of the information of patients including their symptoms of disease, results of various testing such as blood testing and medical history etc. are anticipated by the medical professionals as the input data. Such type of data are handled by medical professionals on the basis of their medical knowledge and experience which they are having for the diagnosis of disease. Occasionally various diseases have few similar symptoms, due to which medical professionals must be allocate arbitrary weights to each one of these inputs and also make patterns, match these patterns with the patterns of the several diseases and at last finally select the match which is closest and finally after the help of discovered knowledge diagnosis the exact disease.

2. **Final or Provisional Diagnosis (FD):** In Final or Provisional Diagnosis phase, generally the initial recommendations and treatments would be to start the treatment according to the identified disease. In such step, normally a physician with good medical knowledge and his or her logic, regular check-ups and also records the results of the findings of continually perceives or tests, and finally on the basis of all these he or she decides the final prescription.

If we see towards the techniques of data mining then we can say it clearly that data mining has various techniques such as: Clustering, Classification, Regression, Association Rules and etc., and it also comprises the algorithms such as: Genetic Algorithm, Genetic Algorithm, Nearest Neighbor method etc., for analysing the plenty amount of unstructured, unprocessed or multi-dimensional data. In the other words we can also say that, data mining also has the abilities for astute analysis of data to excerpt hidden knowledge from the large databases of clinical or the medical data which are collected from various sources like hospitals or medical centers. Such extracted knowledge provides relevant evidence to boost decision support, diagnosis, prevention and treatment in the world of medical science. Moreover, data mining also has the capability to recognize association rules or to set-up relationships between miscellaneous features like: disease symptoms, patient's personal data and etc.

This study investigation endeavour to symbolize the findings of various research works which are explored or published in the field of data mining applications like in the diagnosis, prediction, or treatment of breast cancers.

This research paper is basically organized in to the five different sections. As 1st section inludes introductory, Section 2nd includes some basic concepts related to this as well as some literarture review, Section 3rd presents research methodology, Section 4th includes the proposed model as data mining applications or usages in early diagnosis, treatment and prognosis of various cancers and atlast Section 5th concludes this paper and this section also presents our future works.

Cancers:

Cancer is basically one type of the disease, which happens when the cells growth in any part of human body becomes out-of-control. Or In the other sense, whenever cells in the part of the body divide uncontrollably and which can also damage to the other cells, in such circumstances cancer can occurred. Currently, there are more than hundred types of cancers based on the part of the body where it is appeared, or cells which are affected, have been identified. In present time, one of the main leading cause of the death all over the world is cancer. There are various factors which affects the creation or spreading of cancers includes as: age, marital status, genetics, quality of life, living location and etc.

In most of the cases, cancer makes a collection of tissue in any part of the body which is commonly termed as tumors. Such tumors can grow and can also effect the various organs of the body such as digestive, nervous or circulatory systems. But, when tumor escalate to the other parts of the human body, destroying or invading to other tissues, it is termed as metastasized and whole of this process is termed as metastasis. However, when tumor becomes in this stage, it is really very difficult to treat such tumor. So, one of the most pivotal issues in curing process of tumors and cancers is concerned to the time and stage of diagnosis of that cancer or tumor. Early diagnosis of cancers or tumors increases the chance of their successful curing. Due to this, several re-searchers attempts to create intelligent expert systems so as to to assist the medical professionals for the timely and early diagnosis of cancers.

Generally, in medical science Two types of tumors are identified:
Benign: Benign type of tumors is not much dangerous for the human body and it rarely causes for the human death. In such type, tumor grows in one part (spot) of the body and it has only limited growth.

Malignant: Malignant type of tumors is much more dangerous and it has two types of effects on the human body:

- A cancerous cell along with the uncontrolled progression spread with the invasion lymph system it destroys the healthy tissues. In the other words, it is metastasis to other tissues and also make problem in their general actions or duties.

- It is also the fact that A cancerous cell continually growth and with angiogenesis process makes the new blood vessels to feed to itself. Therefore, it uses the body's blood and can also cause Anemia.

In the present scenario, one of the biggest challenges in the area of cancer treatment is, to identify the most common symptoms which can help for the purpose of earlier diagnosis of cancers. In this context several research studies have already been conducted to excerpt the patterns of cancers and create intelligent or speedy methods for the proper diagnosis of tumors cell in the early stages and suggest the best treatments for the early prevention of the cancer.

Currently most of the medical professionals for identification of the type of cancers (benign breast tumours from malignant) prefer to make surgical biopsy. But

on the other side most of them also believed that, biopsy is very critical task and it must be prevented as much as possible. Therefore, in such a critical situation proposing an intelligent system which can help to medical practitioner to classify the particular type of cancer and also avoid unnecessary surgical biopsy would also be helpful for both medical practitioner and patient. Here we had attempted to cover most of the research works which have been done related to the diagnosis of the cancers and also with applying various computer aided techniques of data mining along with their outcomes. Here for the proper management we will classify the research works which had been done earlier based on the factual points and discuss them.

II. LITERATURE REVIEW

On the basis of the earlier research work done in this field, on the basis of the main goal of the paper, we had categorised them in to the following categories:

Some of the research works, are compared the accuracy of applying various classification techniques for diagnosis of breast cancers, such as:

- Vikas Chaurasia et al. [1] applied Simple Logistic, RBF and RepTree for diagnosis of breast cancer. The accuracy of their classification was 74.5%.

- Wei-pin Chang et al. [2] made a comparative study for predicting breast cancers by decision tree, neural network, genetic algorithm and logistic regression. They concerned on 10 variable/attribute for creating breast cancer classification model. These variables were included: Clump thickness, Bland chromatin, Uni-formity of cell size, Uniformity of cell shape, Bare nuclei, Normal nucleoli, Marginal adhesion, Mitoses, Single epithelial cell size and class variable with two value (benign/malignant). Their experimental results revealed that, decision tree has lowest prediction accuracy and logistic regression model had higher accuracy rate among these applied techniques for predicting breast cancers. Further, genetic algorithm had highest accuracy in the classification of breast cancers and created acceptable classification rules.

- Based on the results of research which is done by Chaurasi and et al. [1], Simple logistic classifier among the other machine learning algorithms with having accuracy of 74.4% and total time taken for building model in 0.62 seconds, is the best algorithm for diagnosis of breast cancers. Further, in this study, researchers used three tests (including Gain Ratio test, Info Gain test and Chi-square test) for recognizing the variables which are important in diagnosis or treatment of breast cancers such as: Tumour size, patients' Age, Degree of malignancy, Menopause, Breast-quad and etc.

- Shweta Kharya [3] made a complete survey about applying different classification techniques for diagnosis of breast cancers. She studied different the performance or accuracy rate of various techniques (including Decision Tree, Bayesian Network, Logistic Regression, Support Vector Machines, Naïve Bayes Cla-ssifier, Association Rule Mining and ANN) for diagnosis of cancers by analysing factors (genes and etc.) or Digital Mammography images classification, Her study was based on the data which are collected from WBCD and SEER datasets. She claimed that, Decision tree with 93.62% accuracy rate of predicting cancers is the best predictor among the concerned techniques and the Bayesian network is the popular technique which is used in medical world for Brest cancer prognosis and diagnosis.

- Senturk et al. [4] applied seven algorithms including KNN, Decision Tree, Naïve bayes, logistic regression, multi-layer perceptron, discriminant analysis and Support Vector Machine for diagnosis of breast cancers. Their experimental results declared that, accuracy of classification made by Support Vector Machine was high than others.

- Ghassem Pour and colleagues [5] made a comparison between a Neural Network classification techniques with Model-based data mining techniques for accuracy of detecting breast cancers. Their experimental results showed that, adding an ensemble oriented approach can improve the results of both techniques. Furthermore, Neural Network approach with ensemble oriented approach had highest accuracy rate of classification in compare with model based data mining techniques.

- Rajesh et al. [6] for classifying patients into either "Carcinoma in situ" (beginning or pre-cancer stage) or "Malignant potential" group, used C4.5 algorithm. They showed that, C4.5 had accuracy ~93% for diagnosis of breast cancers.

- Hota [7] several intelligent techniques such as, ANN (Artificial Neural Network, Unsupervised ANN, Statistical and decision tree based techniques used for classifying data related to breast cancer. In this research work, different models are combined and made ensemble model. Experimental results in this study revealed that, the accuracy rate of ensemble model is better than single individual model.

- Gupta and et al. [8] made a survey with study the several techniques which are used by many researchers for diagnosis and prognosis of breast cancers. Finally they mentioned that, in both cases, for selecting the best technique or algorithm with high degree of accuracy, can be decided after creating several types of models, trying different techniques or algorithms.

- Burke HB et al. [9] compared the prediction accuracy of the TNM staging system1 with that of artificial neural network statistical models. They studied the accuracy of breast cancer prediction based on 5 years and 10 years surveillance data and revealed that, in both case, Artificial Neural Network's prediction was accurate than TNM staging system.

- Ronak Sumbaly et al. [10] used general (types, risk factors, symptoms and treatment) of breast cancers and applied various data mining techniques for diagnosis of breast cancers in the early step. Their results showed that, decision tree have capability to diagnosis breast cancers in the first stages.

- Shrivastava et al. [11] made a review of different classification techniques which have been done for diagnosis of breast cancers. Finally they showed that, Neural Network and decision tree are the most popular techniques which are used by various researchers to create decision rules or predictive models from the breast cancer data.

- Jahanvi Joshi et al. [12] applied various classification and clustering techniques to cre-ate pattern of breast cancer patients. For find-ing the healthy patients, several classifier rules are used. Further, authors claimed that, they used 47 classification algorithms for recognizing healthy people from sick patients. Their experimental results showed that, the results of approximately 13 techniques within those 47 applied techniques were same (24% sick pa-tients and 76% healthy people). These 13 techniques are: Multilayer Perceptron, LMT classifier, Logistic, Classification via Regression, Multi-Class Classifier, GD, SMO, J48, Simple Logistic, AdaBoostM1, Bayes Net and Attribute Selected technique.

- Padmavati et al. [13] for predicting breast cancers used RBF (Radial Basis Function), MLP (Multilayer Perceptron) and Logistic Regression techniques. Their experimental results showed that, RBF has prediction capability of RBF was better that two other techniques. Further, the time taken for prediction by RBF was lesser than other techniques.

- Aboul [14] applied rough set data and ID3 decision tree classifier algorithm for creating classification rules. Their experimental results showed that, the accuracy of classification rules created by rough set was better than ID3 algorithm. Further, the number of classification rules made by rough set algorithm is reduced in compare with ID3 algorithm. In the other words, rough set algorithm had compact number of produced rules.

- Gouda I. Salama et al. [15] compared the accuracy and confusion matrix based on 10-fold cross validation method of different classification techniques including Multi-Layer Perception (MLP), decision tree (J48), Instance Based for K-Nearest neighbour (IBK) and Sequential Minimal Optimization (SMO) for diagnosis of cancers in three different databases of breast cancers (WPBC, WDBC, WBC). Their experimental results showed that, the combination of SMO, MLP, IBK and J48 hast the highest accuracy rate in compare with other techniques (in all of three datasets) for diagnosis of benign breast tumours from malignant.

- Also there are several research works which have attempted to propose a method or approach to recognize benign from malignant breast tumours.

- Hassanien and colleagues [17] studied the applications of rough set theory to analysis the medical data and proposed an approach for creating compact classification rules with applying their proposed simplification algorithm. They claimed that proposed approach had classification accuracy of 98% whereas, accuracy of classification made by decision trees was 85.25. Further, the number of classification rules with applying decision trees and their proposed approach was respectively 428 and 30.

- Einipour [19] combined two methodologies including ACO (Ant Colony Optimization) and Fuzzy System and made an automatically breast cancer diagnosis system named as FUZZY-ACO. The main advantage of the proposed system was high reliability and adequate interpretability in compared with other algorithms. Further the results of comparing the proposed approach with some algorithms such as C4.5, SVM, NN, Naïve Bayes and MLP revealed that, it had accuracy rate higher than other algorithms.

- Raad and colleagues [20] made an approach for classification of breast cancers based on neural network techniques. Further, they developed a tool for automatic detection of breast cancers based on RBF neural network. They proved that accuracy, reliability and efficiency of RBF in compare with MLP technique was better.

- Wen-Jia Kuo at el. [21] proposed a new computer aided diagnosis (CAD) system for classification of breast cancers by using decision tree technique. The main goal was reducing the number of unnecessary biopsies and increasing the diagnosis confidence. They used 24 co-variance texture features for creating decision tree with ability of identifying benign and malignant breast cancers. Accuracy, Positive Predictive value, Negative Predictive value, Sen-

sitivity and Specificity are concerned as objective indices for estimating performance of proposed system in diagnosis of cancers. Authors claimed that, their system which had been made by decision tree had 96% accuracy ra-te, 93.33% Positive Predictive Value, 96.69% Negative Predictive Value, 93.33% sensitivity and 96.67% Specificity.

III. METHODOLOGY

In the field of medical science normally, all the patients based on their stage of cancer lead to follow the similar treatments. While the main query arises here is, can we apply behaviour mining techniques to excerpt the exclusive follow-up of treatments for each of the patient? and Can we apply clustering or classification techniques for the grouping patients on the basis of their tarits, stage of cancer and etc. and also use these groups traits for the identification of high risk person whose curing follow must be diverse from the other patients. Since researcher observed that whenever same treatments and curing process is started for a group of patients those are having same stage of disease, sometimes number of patients not become fit and also their health and extent of cancer becomes larger, whereas for other patients extent and size of cancer becomes smaller. Why? There is no clear-cut answer in medical world for this question. Even sometimes a few of patients even with the stage 2 also even stage 3 of cancer patient without any treatment live along time whereas, on the other hand some other patients whose having same stage of the cancer and continually treated by the physicians, unfortunately no longer time live.

Prognosis problem is also termed as "analysis of survival or lifetime data". It is predicting the occurrence or recurrence of the breast cancer in each individual person. We divide prognosis in two parts [23].On the basis of several result, several attributes are exaggerated on the survivability of cancer. Mainly There are three different types of cancer recurrence:

Local Recurrence: Local recurrence means, breast cancer after sometimes i.e. may be 6 months or more after the complete treatment, and will be back in the same place which it had started before.

Regional Recurrence: Regional recurrence means, when the cancer happens for the second time, it will appear in the lymph nodes near the place that it happened first time.

Distant Recurrence: Distant Recurrence means, after treatment, for a second time when breast cancer appears, it will start in some other part of the body apart from the first one, such as: liver, bone, brain or lungs.

Several experimental results of medical professionals are shown that, approximately most of the person 5 years after the diagnosis of cancer are alive. However, some people live more than 5 years but generally 5-years survival period is used as a standard rate for discussion about prognosis.

Classification: Classification is called as supervise learning. It take some of data (named as training set) which has collection of records and each record contain set of attributes and define one attribute named as class. The main goal of classification is producing a model with capability of predicting the value of class attribute in previously unseen records as accurately as possible. A test set is used for predicting the accuracy of the created model [24]. Some applications of classification in medical diagnosis are: classifying tumor cells, analyzing the effectiveness of treatment and etc.

Several classification algorithms and techniques are proposed such as [25]: Decision Tree Induction (ID3 & C4.5, Hunt's Algorithm and etc.,), Rule-Based Methods, Memory-Based Methods (such as: k-Nearest-Neighbor), Genetic Programming [26], Naïve Bayes [27] and Bayesian Classification [28], Artificial Ne-ural Networks [29], Support Vector Machines (SVMs) [30], Ensemble Methods [31] and etc.

Association Rules: Association Rule is one the most important techniques of data mining. It attempts to extract frequent patterns and interesting relationships between different sets of items [32], and etc.

Regression: Regression same as classification attempts to predict the value of an attribute (variable) based on the value/values of other attributes/variables. The main difference be-tween classification and regression is related to the type of target attribute (variable) that must be predict based on the value of other attributes/variables. The target variable in classification is categorical in nature. Whereas in regression, the target variable is numeric or continuous. Further, in classification, classes are created whereas in regression there is no classes, and all data is divided in various split points and for each split point the amount of "error" is equal to square of differences be-tween amount of actual value and predicted value. The amount of split points error across different variables are compared and minimum split point error is selected as the split point/root node. This process recursively continued. In the other words, the main objectives of regression are:

- Dividing the set of data into two continuous variables then describe the associations or relationships between them.
- Find the value of attributes/variables.
- Predict the value of one attribute/variable based on the value of other attribute/variable.
- Control the accuracy of prediction.

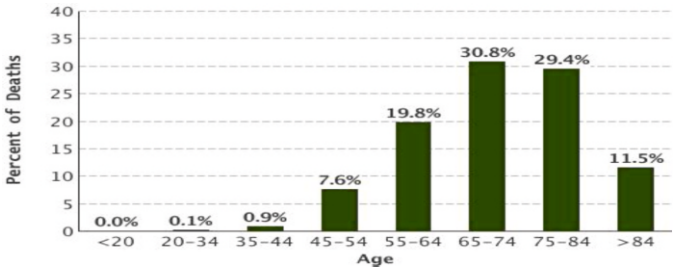
Regression has several applications including: estimation of agricultural data, Geography, marketing, business, Financial Forecasting, medical diagnosis and cancer diagnosis or prognosis, Predicting Laptop Retail Price and etc.

Clustering: Clustering is unsupervised learning and divides the data into groups (call as clusters) based on their similar attributes. All objects into one cluster are similar with each other and dissimilar with objects in other clusters. Clustering is widely used in: science and statistics, pattern recognition, image processing and segmentation, Web applications, DNA analysis in biology, GIS and etc.

Different clustering algorithms are available. Some of these algorithms are: Hierarchical Methods (Divisive Algorithms & Agglomerative Algorithms), Partitioning Methods (Relocation Algorithms, K-medoids Methods, K-means Methods, Probabilistic Clustering, Density-Based Algorithms), Grid-Based Methods, Constraint-Based Clustering, Clustering Algorithms Used in Machine Learning, Scalable Clustering Algorithms, Algorithms For High Dimensional Data (Subspace Clustering, Co-Clustering Techniques, Projection Techniques) and Methods Based on Co-Occurrence of Categorical Data.

IV. PROPOSED SYSTEM

Classification techniques were used for predicting the treatment cost of healthcare services which was increased with rapid growth every year and was becoming a main concern for everyone.



As we know that Clustering is defined as unsupervised learning that occurs by observing only independent variables while supervised learning analysing both independent and dependent variables. It is different from classification which is a supervised learning method. It has no predefined classes. Because of this reason, clustering may be best used for studies of an exploratory nature, mainly if those studies encompass large amount of data, but not very much known about data.

Here, the goal of clustering is descriptive while goal of classification is predictive. The main task of unsupervised learning method means clustering method is to form the clusters from large database on the basis of similarity measure. The goal of clustering is to discover a new set of categories, the new groups are of interest in themselves, and their assessment is intrinsic. In classification tasks, an important part of the assessment is extrinsic. Clustering partitioned the data points based on the similarity measure. Clustering groups data instances into subsets in such a manner that similar instances are grouped together, while different instances belongs to different groups. Clustering approach is used to identify similarities between data points. Each data points within the same cluster are having greater similarity as compare to the data points belongs to other cluster. Clustering of objects is as ancient as the human need for describing the salient characteristics of men and objects and identifying them with a type.

Therefore, it grasp various scientific disciplines: from mathematics and statistics to biology and genetics, each of which uses different terms to describe the topologies formed using this analysis. From biological "taxonomies", to medical "syndromes" and genetic "genotypes" to manufacturing "group technology"— the problem is identical: forming categories of entities and assigning individuals to the proper groups within it. Following are the various clustering algorithms used in healthcare.

Partitional Clustering:

The maximum number of data points in the datasets is 'n'. With the help of 'n' data points the maximum possible number of 'k' clusters is obtained. In order to obtained the 'k' clusters from 'n' data points partitional clustering method is used. In this method, each 'n' data points is relates to one and only 'k' clusters while each 'k' clusters can relates to more than 'n' data points. Partitional clustering algorithms require a user to input k, (which is the number of clusters). Generally, partitional algorithms directly relocate objects to k clusters.

Partitional algorithms are categorized according to how they relocate objects, how they select a cluster centroid (or representative) among objects within a (incomplete) cluster, and how they measure similarities between objects and cluster

centroids. Before we obtained the clusters this method requires to define the required number of cluster which we may have to obtained from datasets. On the basis of similarities between objects and cluster centroids this method is partitioned into two categories. These are K-means and K-Mediods. One of the most popular algorithms of this approach is K-means. First of all it randomly selects k objects and then decomposes these objects into k disjoint groups by iteratively relocating objects based on the similarity between the centroids and objects. In k-means, a cluster centriod is mean value of objects in the cluster. The next algorithm is K-mediods. The major advantage of partitional clustering algorithms is their superior clustering accuracy as compared with hierarchal clustering algorithms that is the result of their global optimization strategy (i.e., the recursive relocations of objects). Another advantage is, partitional algorithms can handle large data sets which hierarchal algorithms cannot (i.e., better scalability) and can more quickly cluster data. In other words we can say that, partitional algorithms are more effective and efficient than hierarchical algorithms. One major drawback to the use of partitional algorithms is that their clustering results depend on the initial cluster centroids to some degree because the centroids are randomly selected.

V. CONCLUSION AND FUTURE WORK

The privacy regarding to patient's confidential information is very important. Such type of privacy may be lost during sharing of data in distributed healthcare environment. Necessary steps must be taken in order to provide proper security so that their confidential information must not be accessed by any unauthorized organizations. But in situations like epidemic, planning better healthcare services for a very large population etc. some confidential data may be provided to the researchers and government organizations or any authorized organizations. In order to achieve better accuracy in the prediction of diseases, improving survivability rate regarding serious death related problems etc. various data mining techniques must be used in combination. This paper also reviewed several research works which are done for diagnosis, treatment or prognosis breast cancers. Based on the results of this study, most of the research works are concerned on comparing the accuracy rate of data mining various algorithms or techniques. Unfortunately, there is no tool that automati¬cally diagnose or prognoses breast cancer. Further, there is no research work which apply personalized features for proposing the best treatment for patients. To achieve medical data of higher quality all the necessary steps must be taken in order to build the better medical information systems which provides accurate information regarding to patients medical history rather than the information regarding to their billing invoices. Because high quality healthcare data is useful for providing better medical services only to the patients but also to the healthcare organizations or any other organizations who are involved in healthcare industry.In the future work, we will attempt to develop a tool with the help of intelligent agents and applying data mining tools with the capability of automatically breast cancer diagnosis and pro¬posing the best treatment.

REFERENCES

[1] Chaurasia V, Pal S. Data Mining Techniques: To Predict and Resolve Breast Cancer Sur-vivability. International Journal of Computer Science and Mobile Computing 2014; 3: 10-22..

[2] Chang PW and Liou MD. Comparison of three Data Mining techniques with Genetic Algorit-hm in analysis of Breast Cancer data. [Onli-ne]. Available: http://edoc.ypu.edu.tw:8080/paper/ha/Other/%E5%BC%B5%E5%81%89%E6%96%8C_comparison%20of%20data%20mining%20in%20breast%20cancer.pdf

[3] Kharya S. Using Data Mining Techniques for Diagnosis and Prognosis of Cancer Disea-se. International Journal of Computer Scien-ce, Engineering and Information Technology (IJCSEIT) 2012; 2: 55-66..

[4] Senturk ZK, Kara R. Breast Cancer Diagnosis via Data mining: Performance Analysis Of Seven Different Algorithms. Computer Science & Engineering: An International Journal (CSEIJ) 2014; 4: 35-46.

[5] Ghassem Pour S, Mc Leod P, Verma B, Maeder A. Comparing Data Mining with Ensemble Classification of Breast Cancer Masses in Digital Mammograms. 2012; http://ceur-ws.org/Vol-941/aih2012_GhassemPour.pdf.

[6] Rajesh K, Anand S. Analysis of SEER Dataset for Breast Cancer Diagnosis using C4.5 Classification Algorithm. International Journal of Advanced Research in Computer and Co-mmunication Engineering 2012; 1: 72-77

[7] Hota HS. Diagnosis of Breast Cancer Using Intelligent Techniques. International Journal of Emerging Science and Engineering (IJESE) 2013; 1: 45-53.

[8] Gupta S, Kumar D, Sharma A. Data Mining Classification Techniques Applied For Breast Cancer Diagnosis and Prognosis. Indian Jou-rnal of Computer Science and Engineering 2011; 2.

[9] Burke HB, Goodman PH, Rosen DB, Henson DE, Weinstein JN, Harrell FE Jr, Marks JR, Winchester DP & Bostwick DG. Artificial Neural Networks Improve the Accuracy of Cancer Survival Prediction. Cancer 1997; Vol. 79,pp.857-862.

[10] Sumbaly R, Vishnusri N, Jeyalatha S. Diagnosis of Breast Cancer using Decision Tree Data Mining Technique. International Journal of Computer Applications, Volume 2014; 98: 16-24..

[11] Shrivastava SS, Sant A, Aharwal RP. An Overview on Data Mining Approach on Breast Cancer data. International Journal of Advanced Computer Research 2013; 3: 256-262.

[12] Joshi J, Doshi R and Patel J. Diagnosis and Prognosis Breast Cancer Using Classification Rules. International Journal of Engineering Research and General Science 2014; 2: 315-323.

[13] Padmavati J. A Comparative study on Bre-ast Cancer Prediction Using RBF and MLP. International Journal of Scientific & Engineering Research 2011; 2: 1-5.

[14] Aboul HE and Jafar AH. Rough set approach for generation of classification rules of Breast cancer data. Journal Informatica, 2004; 15: 23-38.

[15] Salama GI, Abdelhalim MB, Zeid MA. Breast Cancer Diagnosis on Three Different Datasets Using Multi-Classifiers. International Journal of Computer and Information Technology 2012; 1: 36-43.

[16] Sarvestan SA, Safavi AA, Parandeh MN, Salehi M. Predicting Breast Cancer Survivability using data mining techniques. Journal of Software Technology and Engineering (ICSTE) 2010; 2: 227-231.

[17] Yadav R, Khan Z, Saxena H. Chemotherapy Prediction of Cancer Patient by using Data Mining Techniques. International Journal of Computer Applications 2013; 76: 28-31.

[18] Orlando A, Bruno GC, Susana V, Jorge G, Arlindo OL and Jose R. A Data Mining approach for detection of high-risk Breast Cancer groups. Advances in Soft Computing 2010; 74: 43-51.

[19] Einipour A. A Fuzzy-ACO Method for Detect Breast Cancer. Global Journal of Health Sci-ence 2011; 3: 195-199.

[20] Raad A, Kalakech A, Ayache M. Breast Cancer Classification Using Neural Network Approach: MLP and RBF. The 13th International Arab con¬ference on Information Technology ACIT'12 2012; pp. 15-19.

[21] Kuo WJ, Chang RF, Chen DR, Lee CC. Data min¬ing with Decision Trees for Diagnosis of Breast Tumor in Medical Ultrasonic Images. Breast Cancer Research and Treatment 2001; 66: 51-57.

[22] Majali J, Niranjan R, Phatak V and Tadakhe O. International Journal of Computer Science and Information Technologies (IJCSIT) 2014; 5: 6487-6490.

[23] American Cancer Society. When Cancer Comes Back: Cancer Recurrence. http://www.cancer.org/acs/groups/cid/documents/ webcontent/002947-pdf.pdf.

[24] Kumar R, Verma R. Classification Algorithms for Data Mining: A Survey. International Journal of Innovations in Engineering and Technology (IJIET) 2012; 1: 7-14.

[25] Kesavaraj G, Sukumaran S. A Study on Cla-ssification Techniques in Data Mining. 2012; 1: 4th ICCCNT.

[26] Kumar D, Beniwal S. Genetic Algorithm and Programming Based Classification: A Survey. Journal of Theoretical and Applied Information Technology 2013; 54: 48-58.

[27] Vidhya KA and Aghila G. A Survey of Naïve Bayes Machine Learning approach in Text Document Classification. (IJCSIS) International Journal of Computer Science and Information Security 2012; 7: 206-211.

[28] Bielza C and Larranaga P. Discrete Bayesian Network Classifiers: A Survey. ACM-Transaction 2014; 47: 1.

[29] Sood A. Artificial Neural Networks- Growth & Learn: A Survey. International Journal of Soft Computing and Engineering (IJSCE) 2013; 2: 103-104.

[30] Pradhan A. SUPPORT VECTOR MACHINE-A Survey. International Journal of Emerging Te-chnology and Advanced Engineering 2012; 2: 82-85.

[31] Bhatt J and Patel NS. A Survey on One Class Classification using Ensembles Method. IJIRST-International Journal for Innovative Research in Science and Technology 2014; 1: 19-23.

[32] Zhao Q, Bhowmick S. Association Rule Mining: A Survey. Technical Report, CAIS, Nanyang Technological University, Singapore 2003; No. 2003116.